

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2000-003187

(43)Date of publication of application : 07.01.2000

(51)Int. Cl.

G10L 11/04

G10K 15/04

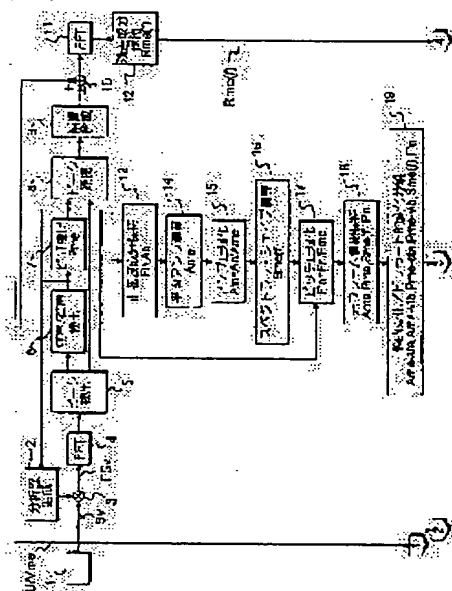
(21)Application number : 10-169046

(71)Applicant : YAMAHA CORP

(22)Date of filing : 16.06.1998

(72)Inventor : KONDO TAKAYASU
XAVIER SERA

(54) METHOD AND DEVICE FOR STORING VOICE FEATURE INFORMATION



(57)Abstract:

PROBLEM TO BE SOLVED: To store precise voice feature information with small storage capacity by obtaining a specified multiplication and a specified difference or ratio related to plural frequencies showing a voice feature.

SOLUTION: An input voice signal segmenting part 3 multiplies an analytic window AW having a period of fixed times of a pitch period generated by an analytic window generation part 2 with an input voice signal Sv. Then, the cut-out part 3 segments the input voice signal Sv in frame to output it to a fast Foulrier transform part (FFT) 4 as a frame voice signal FSv. The signal FSv is analytic processed by the FFT 4, and a local peak value shown by combination between a frequency value and an amplifier value is detected from a frequency spectrum being its output by a peak value detection part 5. Plural frequencies Fk (k is natural number) showing such a voice characteristic are obtained, and a reference frequency Fo is multiplied with respective natural numbers (k), and the difference or the ratio between respective multiplication results $F_0 \times k$ and respective frequencies Fk are obtained to store these difference or the ratio.

Detailed Description of the Invention:

....

[0003]

[Problems to be Solved by the Invention]

In prior-art voice conversion apparatuses, though voice conversion (for example, from male voice to female voice, from female voice to male voice, and the like) is performed, it is simply conversion of voice nature. Therefore, it is impossible to convert voice so that the voice resembles the voice of a particular singer (for example, a professional singer). If there is a function of causing not only voice nature but also the way of singing to resemble those of a particular singer, that is, a function of impersonation, it will be very interesting in a karaoke machine and the like. However, such processing is impossible in prior-art voice conversion apparatuses. Therefore, the inventors provide a voice conversion apparatus capable of causing voice nature to resemble the voice of a target singer.

[0009][2. 4] Step S4

Next, by appropriately selecting and combining attribute data corresponding to a singer (me) trying to do an imitation and target attribute data corresponding to a singer to be imitated, new attribute data (new attribute data = pitch, amplitude and spectrum shape) are obtained.

....

[0010][2. 5] Step S5

Then, based on the subsequently obtained new attribute data, sine wave components of the frame are derived.

[2. 6] Step S6

Then, based on the derived sine wave components and either one of the residual components obtained in step S1 and the residual components of a singer (Target) to be imitated previously memorized (stored), reversed FFT is performed to obtain a conversion voice signals.

[0011][2. 7] Conclusion

By the conversion voice signals resulted from such processing, a reproduced voice becomes just like a singing voice of the other singer (target singer) from a singing voice of a singer trying to do an imitation.

Explanation of Reference numerals:

Figure 1

- 2 analytic window generation
- 5 peak detection
- 6 detection of voiceless/voiced
- 7 pitch detection
- 8 peak linkage
- 9 interpolation/synthesis
- 12 residual component keeping
- 13 sine-wave component keeping
- 14 average amplitude operation
- 15 amplitude normalization
- 16 spectral shape operation
- 17 pitch normalization
- 18 original frame information keeping
- 19 static change/vibrato change separation

Figure.2

- 20 target frame information keeping
- 21 key control/tempo change
- 22 easy synchronization processing
- 23 sine wave component attribute data selection
- 24 attribute data modification
- 24' new sine wave component attribute data
- 25 residual component selection
- 26 generation of sine wave component
- 27 sine wave component modification
- 28 reversed FET
- 31 sequencer
- 32 sound generator
- 34 output

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開2000-3187

(P2000-3187A)

(43) 公開日 平成12年1月7日(2000.1.7)

(51) Int.Cl. ⁷	識別記号	F I	テマコード(参考)
G 1 0 L 11/04		G 1 0 L 9/00	B 5 D 1 0 8
G 1 0 K 15/04	3 0 2	G 1 0 K 15/04	3 0 2 D

審査請求 未請求 請求項の数 5 O L (全 15 頁)

(21) 出願番号	特願平10-169046	(71) 出願人	000004075 ヤマハ株式会社 静岡県浜松市中沢町10番1号
(22) 出願日	平成10年6月16日(1998.6.16)	(72) 発明者	近藤 高康 静岡県浜松市中沢町10番1号 ヤマハ株式会社内
		(72) 発明者	ザビエル セラ スペイン パルセロナ カルデデュー 08440 2-2 ビスカイア19
		(74) 代理人	100098084 弁理士 川▲崎▼ 研二 (外1名) Fターム(参考) 5D108 BF20

(54) 【発明の名称】 音声特徴情報記憶方法および音声特徴情報記憶装置

(57) 【要約】

【課題】 人間の音声の特徴として周波数成分を記憶する際、メモリ容量を削減する。

【解決手段】 人間の音声信号のを周波数分析すると、複数のローカルピークが観察できる。そして、各ローカルピークの周波数は、ピッチ周波数のほぼ整数倍になる。そこで、ピッチ周波数の整数倍の値と各ローカルピークの周波数との差分または割合を求め、得られた差分または割合を記憶することによりメモリ容量を削減する。

【特許請求の範囲】

【請求項1】 音声の特徴を表わす複数の周波数 F_k （但し、 k は自然数）を得る過程と、
基準周波数 F_0 と前記各自然数 k との乗算を行う過程と、
各乗算結果 $F_0 \times k$ と前記各周波数 F_k との差分または割合を求める過程と、
これら差分または割合を記憶することによって前記音声の特徴を記憶することを特徴とする音声特徴情報記憶方法。

【請求項2】 音声の特徴を表わす複数の周波数 F_k （但し、 k は自然数）を得る過程と、
基準周波数 F_0 と前記各自然数 k との乗算を行う過程と、
各乗算結果 $F_0 \times k$ と前記各周波数 F_k との差分または割合を求める過程と、
これら差分または割合のうち所定のスレッシュド値を超えるものを選択する過程と、
これら選択された差分または割合を記憶することによって前記音声の特徴を記憶することを特徴とする音声特徴情報記憶方法。

【請求項3】 音声の特徴を表わす複数の周波数 F_k （但し、 k は自然数）を得る過程と、
基準周波数 F_0 と前記各自然数 k との乗算を行う過程と、
各乗算結果 $F_0 \times k$ と前記各周波数 F_k との差分または割合を求める過程と、
これら差分または割合のうち大きい順に所定数の差分または割合を選択する過程と、
これら選択された差分または割合を記憶することによって前記音声の特徴を記憶することを特徴とする音声特徴情報記憶方法。

【請求項4】 音声の特徴を表わす複数の周波数 F_k （但し、 k は自然数）と、これら周波数 F_k に対応する振幅値 A_k を得る過程と、
前記振幅値 A_k のうち所定値以上であるものに対応する周波数 F_k を選択する過程と、
基準周波数 F_0 と、選択された前記周波数 F_k に対応する各自然数 k との乗算を行う過程と、
各乗算結果 $F_0 \times k$ と前記選択された各周波数 F_k との差分または割合を求める過程と、
これら差分または割合を記憶することによって前記音声の特徴を記憶することを特徴とする音声特徴情報記憶方法。

【請求項5】 請求項1～4の何れかに記載の音声特徴情報記憶方法を実行することを特徴とする音声特徴情報記憶装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、音声信号を生成す

る装置、特にカラオケ装置に用いて好適な音声特徴情報記憶方法および音声特徴情報記憶装置に関する。

【0002】

【従来の技術】入力された音声の周波数特性などを変えて出力する音声変換装置は種々開発されており、例えば、カラオケ装置の中には、歌い手の歌った歌声のピッチを変換して、男性の声を女性の声に、あるいはその逆に変換させるものもある（例えば、特表平8-508581号）。

10 【0003】

【発明が解決しようとする課題】従来の音声変換装置においては、音声の変換（例えば、男声→女声、女声→男声など）は行われるものの、単に声質を変えるだけに止まっていたので、例えば、特定の歌唱者（例えば、プロの歌手）の声に似せるように変換するということはできなかった。また、声質だけでなく、歌い方までも特定の歌唱者に似させるという、ものまねのような機能があれば、カラオケ装置などにおいては大変に面白いが、従来の音声変換装置ではこのような処理は不可能であった。そこで、本発明者らは、声質を目標（ターゲット）とする歌唱者の声に似させることができる音声変換装置を提供することにした。

【0004】しかし、かかる装置においては、音声特徴情報を記憶する必要があるため、膨大な記憶容量が必要である。この発明は上述した事情に鑑みてなされたものであり、僅かな記憶容量で高精度な音声特徴情報を記憶できる音声特徴情報記憶方法および音声特徴情報記憶装置を提供することを目的としている。

【0005】

30 【課題を解決するための手段】上記課題を解決するため請求項1記載の構成にあっては、音声の特徴を表わす複数の周波数 F_k （但し、 k は自然数）を得る過程と、基準周波数 F_0 と前記各自然数 k との乗算を行う過程と、各乗算結果 $F_0 \times k$ と前記各周波数 F_k との差分または割合を求める過程と、これら差分または割合を記憶することによって前記音声の特徴を記憶することを特徴とする。また、請求項2記載の構成にあっては、音声の特徴を表わす複数の周波数 F_k （但し、 k は自然数）を得る過程と、基準周波数 F_0 と前記各自然数 k との乗算を行う過程と、各乗算結果 $F_0 \times k$ と前記各周波数 F_k との差分または割合を求める過程と、これら差分または割合のうち所定のスレッシュド値を超えるものを選択する過程と、これら選択された差分または割合を記憶することによって前記音声の特徴を記憶することを特徴とする。また、請求項3記載の構成にあっては、音声の特徴を表わす複数の周波数 F_k （但し、 k は自然数）を得る過程と、基準周波数 F_0 と前記各自然数 k との乗算を行う過程と、各乗算結果 $F_0 \times k$ と前記各周波数 F_k との差分または割合を求める過程と、これら差分または割合のうち大きい順に所定数の差分または割合を選択する過

程と、これら選択された差分または割合を記憶することによって前記音声の特徴を記憶することを特徴とする。また、請求項4記載の構成にあっては、音声の特徴を表わす複数の周波数 F_k （但し、 k は自然数）と、これら周波数 F_k に対応する振幅値 A_k を得る過程と、前記振幅値 A_k のうち所定値以上であるものに対応する周波数 F_k を選択する過程と、基準周波数 F_0 と、選択された前記周波数 F_k に対応する各自然数 k との乗算を行う過程と、各乗算結果 $F_0 \times k$ と前記選択された各周波数 F_k との差分または割合を求める過程と、これら差分または割合を記憶することによって前記音声の特徴を記憶することを特徴とする。また、請求項1～4の何れかに記載の音声特徴情報記憶方法を実行することを特徴とする。

【0006】

【発明の実施の形態】[1] 実施形態の概要処理

[2] 実施形態の概要処理

次に図面を参照して本発明の好適な実施形態について説明する。始めに、実施形態の概要処理について説明する。

[2.1] ステップS1

まず、ものまねをしようとする歌唱者（me）の音声（入力音声信号）をリアルタイムでFFT（Fast Fourier Transform）する過程を含むSMS（Spectral Modeling Synthesis）分析を行い、フレーム単位で正弦波成分（Sine成分）を抽出するとともに、入力音声信号及び正弦波成分からフレーム単位で残差成分（Residual成分）を生成する。これと並行して入力音声信号が無声音（含む無音）か否かを判別し、無声音である場合には、以下のステップS2～ステップS6の処理は行わず、入力音声信号をそのまま出力することとなる。この場合において、SMS分析としては、前回のフレームにおけるピッチに応じて分析窓幅を変更するピッチ同期分析を採用している。

【0007】[2.2] ステップS2

次に入力音声信号が有声音である場合には、抽出した正弦波成分からさらに元属性（Attribute）データであるピッチ（Pitch）、アンプ（Amplitude）及びスペクトラル・シェイプ（Spectral Shape）を抽出する。さらに抽出したピッチ及びアンプについては、ビブラート成分及びビブラート成分以外の他の成分に分離する。

【0008】[2.3] ステップS3

予め記憶（保存）してあるものまねの対象（Target）となる歌唱者の属性データ（ターゲット属性データ＝ピッチ、アンプ及びスペクトラル・シェイプ）から、ものまねをしようとする歌唱者（me）の入力音声信号のフレームに対応するフレームのターゲット属性データ（＝ピッチ、アンプ及びスペクトラル・シェイプ）を取り出す。この場合において、ものまねをしようとする歌唱者（me）の入力音声信号のフレームに対応するフレームのタ

ーゲット属性データが存在しない場合には、後に詳述するように、予め定めたイーザシンクロナイゼーション規則（Easy Synchronization Rule）に従って、ターゲット属性データを生成し、同様の処理を行う。

【0009】[2.4] ステップS4

次にものまねをしようとする歌唱者（me）に対応する元属性データ及びものまねの対象となる歌唱者に対応するターゲット属性データを適宜選択して組み合わせることにより、新しい属性データ（新属性データ＝ピッチ、アンプ及びスペクトラル・シェイプ）を得る。なお、ものまねではなく、単なる音声変換として用いる場合には、元属性データ及びターゲット属性データの加算平均として新属性データを得るなどの元属性データ及びターゲット属性データの双方に基づいて計算により新属性データを得るようにすることも可能である。

【0010】[2.5] ステップS5

つづいて得られた新属性データに基づいて、当該フレームの正弦波成分を求める。

[2.6] ステップS6

そして求めた正弦波成分と、ステップS1で求めた残差成分あるいは予め記憶（保存）してあるものまねの対象（Target）となる歌唱者の残差成分のいずれか一方と、に基づいて逆FFTを行い、変換音声信号を得る。

【0011】[2.7] まとめ

これらの処理の結果得られる変換音声信号によれば、再生される音声は、物まねをしようとする歌唱者の歌声が、あたかも、別の歌唱者（ターゲットの歌唱者）が歌った歌声のようになる。

【0012】[3] 実施形態の詳細構成

図1及び図2に、実施形態の詳細構成図を示す。なお、本実施形態は、本発明による音声変換装置（音声変換方法）をカラオケ装置に適用し、ものまねを行うことができるカラオケ装置として構成した場合の例である。

【0013】図1において、マイク1は、ものまねをしようとする歌唱者（me）の声を収集し、入力音声信号 S_v として入力音声信号切出部3に出力する。これと並行して、分析窓生成部2は、前回のフレームで検出したピッチの周期の固定倍（例えば、3.5倍など）の周期を有する分析窓（例えば、ハミング窓）AWを生成し、入力音声信号切出部3に出力する。なお、初期状態あるいは前回のフレームが無声音（含む無音）の場合には、予め設定した固定周期の分析窓を分析窓AWとして入力音声信号切出部3に出力する。

【0014】これらにより入力音声信号切出部3は、入力された分析窓AWと入力音声信号 S_v とを掛け合わせ、入力音声信号 S_v をフレーム単位で切り出し、フレーム音声信号 $F S_v$ として高速フーリエ変換部4に出力される。

【0015】より具体的には、入力音声信号 S_v とフレームとの関係は、図3に示すようになっており、各フレ

ームFLは、前のフレームFLと一部重なるように設定されている。そして、高速フーリエ変換部4においてフレーム音声信号FSvは、解析処理されるとともに、図4に示すように、高速フーリエ変換部4の出力である周波数スペクトルからピーク検出部5によりローカルピークが検出される。

【0016】より具体的には、図4に示すような周波数スペクトルに対して、×印を付けたローカルピークを検出する。このローカルピークは、周波数値とアンプ（振幅）値の組み合わせとして表される。すなわち、図4に示すように、(F0, A0)、(F1, A1)、(F2, A2)、……、(FN, AN)というように各フレームについてローカルピークが検出され、表されることとなる。

【0017】そして、図3に模式的に示すように、各フレーム毎に一組（以下、ローカルピーク組という。）として無声／有声検出部6及びピーク連携部8に出力される。無声／有声検出部6は、入力されたフレーム毎のローカルピークに基づいて、高周波成分の大きさに応じて無声であることを検出（‘t’、‘k’等）し、無声／有声検出信号U/Vmeをピッチ検出部7、イーージンクロナイゼーション処理部22及びクロスフェーダ部30に出力する。あるいは、時間軸上で単位時間あたりの零クロス数に応じて無声であることを検出（‘s’等）し、元無声／有声検出信号U/Vmeをピッチ検出部7、イーージンクロナイゼーション処理部22及びクロスフェーダ部30に出力する。

【0018】さらに無声／有声検出部6は、入力されたフレームについて無声であると検出されなかった場合には、入力されたローカルピーク組をそのまま、ピッチ検出部7に出力する。ピッチ検出部7は、入力されたローカルピーク組に基づいて、当該ローカルピーク組が対応するフレームのピッチPmeを検出する。

【0019】より具体的なフレームのピッチPmeの検出方法としては、例えば、Maher, R.C. and J.W. Beauchamp: "Fundamental Frequency Estimation of Musical Signal using a two-way Mismatch Procedure" (Journal of Acoustical Society of America 95(4):2254-2263) に開示されているような方法で行う。

【0020】次に、ピーク検出部5から出力されたローカルピーク組は、ピーク連携部8において、前後のフレームについて連携が判断され、連携すると認められるローカルピークについては、一連のデータ列となるようにローカルピークをつなげる連携処理がなされる。

【0021】ここで、この連携処理について、図5を参照して説明する。今、図5(A)に示すようなローカルピークが前回のフレームにおいて検出され、図5(B)に示すようなローカルピークが今回のフレームにおいて検出されたとする。

【0022】この場合、ピーク連携部8は、前回のフレ

ームで検出された各ローカルピーク(F0, A0)、(F1, A1)、(F2, A2)、……、(FN, AN)に対応するローカルピークが今回のフレームでも検出されたか否かを調べる。対応するローカルピークがあるか否かの判断は、前回のフレームで検出されたローカルピークの周波数を中心にした所定範囲内に今回のフレームのローカルピークが検出されるか否かによって行われる。

【0023】より具体的には、図5の例では、ローカルピーク(F0, A0)、(F1, A1)、(F2, A2)……については、対応するローカルピークが検出されているが、ローカルピーク(FK, AK)については(図5(A)参照)、対応するローカルピーク(図5(B)参照)は検出されていない。

【0024】ピーク連携部8は、対応するローカルピークを検出した場合は、それらを時系列順に繋げて一組のデータ列として出力する。なお、対応するローカルピークが検出されない場合は、当該フレームについての対応ローカルピークは無しということを示すデータに置き換える。ここで、図6は、複数のフレームにわたるローカルピークの周波数F0及び周波数F1の変化の一例を示している。

【0025】このような変化は、アンプ（振幅）A0、A1、A2、……についても同様に認められる。この場合、ピーク連携部8から出力されるデータ列は、フレームの間隔おきに出力される離散的な値である。なお、ピーク連携部8から出力されるピーク値を、以後において、確定成分という。これは、元の信号（すなわち、音声信号Sv）のうち正弦波の要素として確定的に置き換えられる成分という意味である。また、置き換えられた各正弦波（厳密には、正弦波のパラメータである周波数及びアンプ（振幅））の各々については、正弦波成分と呼ぶことにする。

【0026】次に、補間合成部9は、ピーク連携部8から出力される確定成分について補間処理を行い、補間後の確定成分に基づいていわゆるオシレータ方式で波形合成を行う。この場合の補間の間隔は、後述する出力部34が出力する最終出力信号のサンプリングレート（例えば、44.1 KHz）に対応した間隔で行われる。前述した図6に示す実線は、正弦波成分の周波数F0、F1について補間処理が行われた場合のイメージを示している。

【0027】[3.1] 補間合成部の構成

ここで、補間合成部9の構成を図7に示す。補間合成部9は、複数の部分波形発生部9aを備えて構成されており、各部分波形発生部9aは、指定された正弦波成分の周波数(F0, F1, …)およびアンプ（振幅）に応じた正弦波を発生する。ただし、本第1実施形態における正弦波成分(F0, A0)、(F1, A1)、(F2, A2)、……は、各々補間の間隔に従って時時刻々変化

10

20

30

40

50

していくものであるから、各部分波形発生部9aから出力される波形は、その変化に従った波形になる。

【0028】すなわち、ピーク連携部8からは正弦波成分(F0、A0)、(F1、A1)、(F2、A2)、……が順次出力され、各正弦波成分の各々について補間処理が行われるから、各部分波形発生部9aは、所定の周波数領域内で周波数と振幅が変動する波形を出力する。そして、各部分波形発生部9aから出力された波形は、加算部9bにおいて加算合成される。したがって、補間合成部9の出力信号は、入力音声信号Svから確定成分を抽出した正弦波成分合成信号SSkになる。

【0029】[3. 1. 1] 補間合成部9のデータ構成

ここで、補間合成部9におけるデータ構成について説明する。補間合成部9は各部分成分について周波数とアンプのペアをN+1組有しており、高い精度でこれらのデータを記憶しようとする、膨大なメモリ容量が必要になる。一方、データの有効桁を少なくする等の手法により精度を下げると、音声信号の忠実度も下がる。

【0030】一方、人間の音声信号の性質として、各部分成分の周波数は、ピッチ周波数のほぼ整数倍になる。そこで、この性質を利用して、ピッチ周波数の整数倍の値と各部分成分の周波数との相違に着目すれば、少ないメモリ容量で忠実な再生が可能であると考えられる。具体的には、以下に述べる何れかの方法、またはこれらの組み合わせを採用すると好適である。

【0031】(1) 差分を記憶する方法

周波数Fk(但しk=0~N)は、以下のように表わすことができる。

$$F_k = F_0 \times k + dF_k$$

ここで、F0×kは、ピッチ周波数の整数倍の値であり、dFkは、この整数倍の値と実際の周波数Fkとの差分値である。実際に周波数Fkを記憶せず、差分値dFkを記憶しておくことにより、周波数Fkを記憶するためのメモリ容量を削減することができる。但し、差分値dFkの取りうる可能性のある値の最小値および最大値は、値「k」に比例して増加する。

【0032】(2) 比率を記憶する方法

周波数Fk(但しk=0~N)は、以下の式によっても表わすことができる。

$$F_k = F_0 \times k \times rF_k$$

ここで、rFkは、ピッチ周波数の整数倍の値F0×kと実際の周波数Fkとの比率である。そこで、実際に周波数Fkを記憶せず、比率rFkを記憶しておくことにより、周波数Fkを記憶するためのメモリ容量を削減することができる。この比率rFkは、値「k」に関係なくほぼ一定である点で上述した差分値dFkよりも扱いが容易である。

【0033】ここで、上記各周波数Fkの採りうる範囲は50Hz~10kHz程度確保しておけば充分である。

一方、比率rFkの範囲は個人差があるが、本発明者らが観測したところによれば、1音程(100セント)程度確保しておけば大部分の人の音声を忠実に再現することができる。また、比率rFkの精度は1セント程度確保しておけば充分である。

【0034】(3) 比率を対数値で記憶する方法

上記比率rFkを記憶する際、これを対数値に変換しておくと、メモリ容量を一層削減することができる。この対数値cFkとして「セント」を用いるとすれば、対数値cFkは下式により求まる。

$$cF_k = 1200 \times \log_2(rF_k)$$

具体的には、対数値cFkによって+/-100セントの範囲を1セントの精度で表現するためには、対数値cFkを8ビットで記憶させるとよい。

【0035】(4) 一部の周波数Fkの記憶を省略する方法

この方法は、上記方法(1)~(3)の何れかの方法と組み合わせる方法である。上記方法(1)~(3)においては、「(N+1)個」の差分値dFk、比率rFkまたは対数値cFkが必要であると考えられる。しかし、本発明者らの実験によれば、一部の周波数Fkについてはピッチ周波数F0の整数倍であると仮定したとしても音質上の劣化が少ないことが判明した。かかる部分成分においては、周波数Fkとしてピッチ周波数F0の整数倍の値を用いることができ、対数値cFk等を記憶する必要がなくなる。

【0036】方法(3)、(4)を組み合わせることを想定すると、周波数Fkを忠実に再現すべき部分成分は、以下の方法(4.1)~(4.3)を採用して決定することができる。なお、方法(4.1)~(4.3)は単独で用いてもよく、組み合わせで用いてもよい。

(4.1)再現数M(但しM<N+1)を予め決定しておき、対数値cFkの大きい順にM個の部分成分を選択する。

(4.2)対数値cFkに対してスレッシュホールド値を決定しておき、対数値cFkが該スレッシュホールド値を超えた部分成分を選択する。

(4.3)アンプの大きさが所定の条件(例えば最大のAkに対して-30dBよりも大きい値)を満たす部分成分を選択する。

【0037】以上のように選択された部分成分に係る周波数情報をM個記憶する場合、メモリの所定の領域に値Mを記憶し、何番目の成分に対応するかを示す値kと、周波数を特定するための情報(上記対数値cFk等)とをM組記憶するとよい。かかる方法は、部分成分数N+1よりも再現数Mがかなり小さい場合に特に有効である。

【0038】(5) アンプAkの記憶方法

この方法は、上記方法(4)またはこれと方法(1)~(3)とを組み合わせる方法である。本実施形態においては、上述したように各部分成分に対してアンプA

kが記憶される。各アンプAkに対して1バイト(8ビット)を割り当て、データの精度を1dBにすると、アンプAkは256dBのダイナミックレンジを有することになる。しかし、実際にはこのように広いダイナミックレンジは不要であり、128dB程度確保できれば充分である。

【0039】通常のコンピュータにおいては、データ長を8ビット単位で設定するため、アンプAkに7ビットを割り当てて128dBのダイナミックレンジを確保すると、1ビット余剰が生じることになる。そこで、この1ビットにおいて、周波数を特定するための情報(上記対数値cfk等)が存在するか否かを示すことにする。以下、この情報をフラグxkという

【0040】そうすると、何番目の成分に対応するかを示す値kを記憶するために独立の記憶領域を設ける必要がなくなるため、メモリ容量を一層削減することができる。なお、アンプAkのデシベル値の表現の仕方についても種々の態様が考えられる。例えば、アンプAkのうち最大値を0dBとして、0～-127dBの範囲で表現してもよい。また、最大値が0dBを超えた値を持つ場合、あるいは、必要なダイナミックレンジが狭く高い分解能が望まれる場合は、下式によりアンプnAkを求め、アンプAkに代えてアンプnAkを記憶してもよい。

$$nAk = \alpha \cdot (Ak + \beta)$$

【0041】すなわち、上式においてアンプnAkが0～127の範囲に収まるように α 又は β を決定するとよい。例として必要なダイナミックレンジが+20dB～-40dBであった場合、 $\alpha = -127/60$ 、 $\beta = -20$ とすると、アンプAkが20dBの時は $nAk = 0$ 、アンプAkが-40dBの時は $nAk = 127$ となり、記憶エリアを有効に利用できる。 α および β の値は予め決定しておき固定値にしてもよく、状況に応じて変化させたい場合は一方、または双方の値もデータ列に加える等の措置により、可変にしてもよい。但し、上記の方法において0～127の範囲を超えたデータが存在した場合は、その範囲に入るように、0以下の値は0に、127以上の値は127に揃えることは必要であろう。

【0042】(6) まとめ

上記方法(3)、(4)および(5)を総合すると、周波数Fkを忠実に再現すべき正弦波成分においては、アンプAk(7ビット)と、フラグxk(1ビット)と、対数値cfk(8ビット)とによって合計16ビットのメモリ容量が必要になる。また、周波数Fkをピッチ周波数F0の整数倍に近似して良い場合は、アンプAk(7ビット)と、フラグxk(1ビット)とによって合計8ビットのメモリ容量が必要になる。一般的な浮動小数点データは32ビット長が必要であるから、アンプAkおよび周波数Fkを合わせると、正弦波成分あたり64ビットが必要になる。本実施形態においては、これを8～16ビットに削減できるから、所要メモリ容量は1/8～1/4

程度に削減することが可能である。

【0043】[3. 2] 残差成分検出部の動作

次に、残差成分検出部10は、補間合成部9から出力された正弦波成分合成信号SSと入力音声信号Svとの偏差である残差成分信号SRD(時間波形)を生成する。この残差成分信号SRDは、音声に含まれる無声成分を多く含む。一方、前述の正弦波成分合成信号SSは有声成分に対応するものである。ところで、目標(Target)となる歌唱者の声に似せるには、有声音についてだけ処理を行えば、無声音については処理を施す必要はあまりない。そこで、本実施形態においては、有声音成分に対応する確定成分について音声変換処理を行うようにしている。より具体的には、残差成分信号SRDについては、高速フーリエ変換部11で、周波数波形に変換し、得られた残差成分信号(周波数波形)を $R_{me}(f)$ として残差成分保持部12に保持しておく。

【0044】[3. 3] 平均アンプ演算部の動作

一方、図8(A)に示すように、ピーク検出部5からピーク連携部8を介して出力された正弦波成分(F0、A0)、(F1、A1)、(F2、A2)、……、(FN-1、A(N-1))のN個の正弦波成分(以下、これらをまとめてFn、Anと表記する。n=0～(N-1))は、正弦波成分保持部13に保持されるとともに、アンプAnは平均アンプ演算部14に入力され、各フレーム毎に次式により平均アンプAmeが算出される。

$$A_{me} = \sum (A_n) / N$$

【0045】[3. 4] アンプ正規化部の動作

次にアンプ正規化部15において、次式により各アンプAnを平均アンプAmeで正規化し、正規化アンプA'nを求める。

$$A'_n = A_n / A_{me}$$

【0046】[3. 5] スペクトラル・シェイプ演算部の動作

そして、スペクトラル・シェイプ演算部16において、図8(B)に示すように、周波数Fn及び正規化アンプA'nにより得られる正弦波成分(Fn、A'n)をブレイクポイントとするエンベロープ(包絡線)をスペクトラル・シェイプSme(f)として生成する。この場合において、二つのブレイクポイント間の周波数におけるアンプの値は、当該二つのブレイクポイントを、例えば、直線補間することにより算出する。なお、補間の方法は直線補間に限られるものではない。

【0047】[3. 6] ピッチ正規化部の動作

続いてピッチ正規化部17においては、各周波数Fnをピッチ検出部7において検出したピッチPmeで正規化し、正規化周波数F'nを求める。 $F'_n = F_n / P_{me}$ これらの結果、元フレーム情報保持部18は、入力音声信号Swに含まれる正弦波成分に対応する元属性データである平均アンプAme、ピッチPme、スペクトラル・シェイプSme(f)、正規化周波数F'nを保持することと

なる。

【0048】なお、この場合において、正規化周波数 $F'n$ は、倍音列の周波数の相対値を表しており、もし、フレームの倍音構造を完全倍音構造であるとして取り扱うならば、保持する必要はない。この場合において、男声/女声変換を行おうとしている場合には、この段階において、男声→女声変換を行う場合には、ピッチをオクターブ上げ、女声→男声変換を行う場合にはピッチをオクターブ下げる男声/女声ピッチ制御処理を行うようにするのが好ましい。

【0049】つづいて、元フレーム情報保持部18に保持している元属性データのうち、平均アンプ A_{me} およびピッチ P_{me} については、さらに静的変化/ビブラートの成分分離部19により、フィルタリング処理などを行って、静的変化成分とビブラート変化的成分とに分離して保持する。なお、さらにビブラート変化的成分からより高周波変化成分であるジッタ変化的成分を分離するように構成することも可能である。

【0050】より具体的には、平均アンプ A_{me} を平均アンプ静的成分 A_{me-sta} 及び平均アンプビブラートの成分 A_{me-vib} とに分離して保持する。また、ピッチ P_{me} をピッチ静的成分 P_{me-sta} 及びピッチビブラートの成分 P_{me-vib} とに分離して保持する。

【0051】これらの結果、対応するフレームの元フレーム情報データ I_{NFme} は、図8(C)に示すように、入力音声信号 S_v の正弦波成分に対応する元属性データである平均アンプ静的成分 A_{me-sta} 、平均アンプビブラートの成分 A_{me-vib} 、ピッチ静的成分 P_{me-sta} 、ピッチビブラートの成分 P_{me-vib} 、スペクトラル・シェイプ $S_{me}(f)$ 、正規化周波数 $F'n$ 及び残差成分 $R_{me}(f)$ の形で保持されることとなる。

【0052】一方、ものまねの対象 (target) となる歌唱者に対応するターゲット属性データから構成されるターゲットフレーム情報データ I_{NFtar} は、予め分析されてターゲットフレーム情報保持部20を構成するハードディスクなどに予め保持されている。この場合において、ターゲットフレーム情報データ I_{NFtar} のうち、正弦波成分に対応するターゲット属性データとしては、平均アンプ静的成分 $A_{tar-sta}$ 、平均アンプビブラートの成分 $A_{tar-vib}$ 、ピッチ静的成分 $P_{tar-sta}$ 、ピッチビブラートの成分 $P_{tar-vib}$ 、スペクトラル・シェイプ $S_{tar}(f)$ がある。

【0053】また、ターゲットフレーム情報データ I_{NFtar} のうち、残差成分に対応するターゲット属性データとしては、残差成分 $R_{tar}(f)$ がある。

【0054】[3.7] キーコントロール/テンポチェンジ部の動作

次にキーコントロール/テンポチェンジ部21は、シーケンサ31からの同期信号 S_{SYNC} に基づいて、ターゲットフレーム情報保持部20から同期信号 S_{SYNC} に対応す

るフレームのターゲットフレーム情報 I_{NFtar} の読出処理及び読み出したターゲットフレーム情報データ I_{NFtar} を構成するターゲット属性データの補正処理を行うとともに、読み出したターゲットフレーム情報 I_{NFtar} および当該フレームが無声であるか有声であるかを表すターゲット無声/有声検出信号 U/V_{tar} を出力する。

【0055】より具体的には、キーコントロール/テンポチェンジ部21の図示しないキーコントロールユニットは、カラオケ装置のキーを基準より上げ下げした場合、ターゲット属性データであるピッチ静的成分 $P_{tar-sta}$ 及びピッチビブラートの成分 $P_{tar-vib}$ についても、同じだけ上げ下げする補正処理を行う。例えば、50 [cent] だけキーを上げた場合には、ピッチ静的成分 $P_{tar-sta}$ 及びピッチビブラートの成分 $P_{tar-vib}$ についても50 [cent] だけ上げなければならない。

【0056】また、キーコントロール/テンポチェンジ部21の図示しないテンポチェンジユニットは、カラオケ装置のテンポを上げ下げした場合には、変更後のテンポに相当するタイミングで、ターゲットフレーム情報データ I_{NFtar} の読み出し処理を行う必要がある。この場合において、必要なフレームに対応するタイミングに相当するターゲットフレーム情報データ I_{NFtar} が存在しない場合には、当該必要なフレームのタイミングの前後のタイミングに存在する二つのフレームのターゲットフレーム情報データ I_{NFtar} を読み出し、これら二つのターゲットフレーム情報データ I_{NFtar} により補間処理を行い、当該必要なタイミングにおけるフレームのターゲットフレーム情報データ I_{NFtar} 、ひいては、ターゲット属性データを生成する。

【0057】この場合において、ビブラートの成分 (平均アンプビブラートの成分 $A_{tar-vib}$ 及びピッチビブラートの成分 $P_{tar-vib}$) に関しては、そのままでは、ビブラートの周期自体が変化してしまい、不適当であるので、周期が変動しないような補間処理を行う必要がある。又は、ターゲット属性データとして、ビブラートの軌跡そのものを表すデータではなく、ビブラート周期及びビブラート深さのパラメータを保持し、実際の軌跡を演算により求めるようにすれば、この不具合を回避することができる。

【0058】[3.8] イージーシンクロナイゼーション処理部の動作

次にイージーシンクロナイゼーション処理部22は、ものまねをしようとする歌唱者のフレーム (以下、元フレームという。) に元フレーム情報データ I_{NFme} が存在するにもかかわらず、対応するものまねの対象となる歌唱者のフレーム (以下、ターゲットフレームという。) にターゲットフレーム情報データ I_{NFtar} が存在しない場合には、当該ターゲットフレームの前後方向に存在するフレームのターゲットフレーム情報データ I_{NFtar}

rを当該ターゲットフレームのターゲットフレーム情報データINF tarとするイージーシンクロナイゼーション処理を行う。

【0059】そして、イージーシンクロナイゼーション処理部22は、後述する置換済ターゲットフレーム情報データINF tar-syncに含まれるターゲット属性データのうち正弦波成分に関するターゲット属性データ（平均アンブ静的成分A tar-sync-sta、平均アンブビブラートの成分A tar-sync-vib、ピッチ静的成分P tar-sync-sta、ピッチビブラートの成分P tar-sync-vib及びスペクトラル・シェイプS tar-sync(f)）を正弦波成分属性データ選択部23に出力する。

【0060】また、イージーシンクロナイゼーション処理部22は、後述する置換済ターゲットフレーム情報データINF tar-syncに含まれるターゲット属性データのうち残差成分に関するターゲット属性データ（残差成分R tar-sync(f)）を残差成分選択部25に出力する。

【0061】このイージーシンクロナイゼーション処理部22における処理においても、ビブラートの成分（平均アンブビブラートの成分A tar-vib及びピッチビブラートの成分P tar-vib）に関しては、そのままでは、ビブラートの周期自体が変化してしまい、不適当であるので、周期が変動しないような補間処理を行う必要がある。又は、ターゲット属性データとして、ビブラートの軌跡そのものを表すデータではなく、ビブラート周期及びビブラート深さのパラメータを保持し、実際の軌跡を演算により求めるようにすれば、この不具合を回避することができる。

【0062】[3. 8. 1] イージーシンクロナイゼーション処理の詳細

ここで、図9及び図10を参照してイージーシンクロナイゼーション処理について詳細に説明する。図9は、イージーシンクロナイゼーション処理のタイミングチャートであり、図10はイージーシンクロナイゼーション処理フローチャートである。

【0063】まず、イージーシンクロナイゼーション処理部22は、シンクロナイゼーション処理の方法を表すシンクロナイゼーションモード="0"とする（ステップS11）。このシンクロナイゼーションモード="0"は、元フレームに対応するターゲットフレームにターゲットフレーム情報データINF tarが存在する通常処理の場合に相当する。

【0064】そしてあるタイミングtにおける元無声/有声検出信号U/V_{me}(t)が無声(U)から有声(V)に変化したか否かを判別する（ステップS12）。例えば、図9に示すように、タイミングt=t1においては、元無声/有声検出信号U/V_{me}(t)が無声(U)から有声(V)に変化している。

【0065】ステップS12の判別において、元無声/有声検出信号U/V_{me}(t)が無声(U)から有声(V)

に変化している場合には（ステップS12: Yes）、タイミングtの前のタイミングt-1における元無声/有声検出信号U/V_{me}(t-1)が無声(U)かつターゲット無声/有声検出信号U/V_{tar}(t-1)が無声(U)であるか否かを判別する（ステップS18）。例えば、図9に示すように、タイミングt=t0(=t1-1)においては、元無声/有声検出信号U/V_{me}(t-1)が無声(U)かつターゲット無声/有声検出信号U/V_{tar}(t-1)が無声(U)となっている。

10 【0066】ステップS18の判別において、元無声/有声検出信号U/V_{me}(t-1)が無声(U)かつターゲット無声/有声検出信号U/V_{tar}(t-1)が無声(U)となっている場合には（ステップS18: Yes）、当該ターゲットフレームには、ターゲットフレーム情報データINF tarが存在しないので、シンクロナイゼーションモード="1"とし、置換用のターゲットフレーム情報データINF holdを当該ターゲットフレームの後方向（Backward）に存在するフレームのターゲットフレーム情報とする。

20 【0067】例えば、図9に示すように、タイミングt=t1~t2のターゲットフレームには、ターゲットフレーム情報データINF tarが存在しないので、シンクロナイゼーションモード="1"とし、置換用ターゲットフレーム情報データINF holdを当該ターゲットフレームの後方向に存在するフレーム（すなわち、タイミングt=t2~t3に存在するフレーム）のターゲットフレーム情報データbackwardとする。

30 【0068】そして、処理をステップS15に移行し、シンクロナイゼーションモード="0"であるか否かを判別する（ステップS15）。ステップS15の判別において、シンクロナイゼーションモード="0"である場合には、タイミングtにおける元フレームに対応するターゲットフレームにターゲットフレーム情報データINF tar(t)が存在する場合、すなわち、通常処理であるので、置換済ターゲットフレーム情報データINF tar-syncをターゲットフレーム情報データINF tar(t)とする。

40 【0069】INF tar-sync=INF tar(t)
例えば、図9に示すようにタイミングt=t2~t3のターゲットフレームには、ターゲットフレーム情報データINF tarが存在するので、
INF tar-sync=INF tar(t)
とする。

50 【0070】この場合において、以降の処理に用いられる置換済ターゲットフレーム情報データINF tar-syncに含まれるターゲット属性データ（平均アンブ静的成分A tar-sync-sta、平均アンブビブラートの成分A tar-sync-vib、ピッチ静的成分P tar-sync-sta、ピッチビブラートの成分P tar-sync-vib、スペクトラル・シェイプS tar-sync(f)及び残差成分R tar-sync(f)）は実質的に

は、以下の内容となる（ステップS16）。

【0071】

A tar-sync-sta = A tar-sta

A tar-sync-vib = A tar-vib

P tar-sync-sta = P tar-sta

P tar-sync-vib = P tar-vib

S tar-sync(f) = S tar(f)

R tar-sync(f) = R tar(f)

【0072】ステップS15の判別において、シンクロ
ナイゼーションモード = “1” またはシンクロナイゼー
ションモード = “2” である場合には、タイミングtに
おける元フレームに対応するターゲットフレームにター
ゲットフレーム情報データINF tar(t)が存在しない場
合であるので、置換済ターゲットフレーム情報データI
NF tar-syncを置換用ターゲットフレーム情報データI
NF holdとする。

【0073】INF tar-sync = INF hold

例えば、図9に示すように、タイミングt = t1~t2の
ターゲットフレームには、ターゲットフレーム情報デー
タINF tarが存在せず、シンクロナイゼーションモー
ド = “1” となるが、タイミングt = t2~t3のター
ゲットフレームには、ターゲットフレーム情報データI
NF tarが存在するので、置換済ターゲットフレーム情報
データINF tar-syncをタイミングt = t2~t3のター
ゲットフレームのターゲットフレーム情報データである
置換用ターゲットフレーム情報データINF holdとする
処理P1を行い、以降の処理に用いられる置換済ター
ゲットフレーム情報データINF tar-syncに含まれるター
ゲット属性データは、平均アンプ静的成分A tar-sync-s
ta、平均アンプビブラートの成分A tar sync vib、ピ
ッチ静的成分P tar-sync-sta、ピッチビブラートの成分P
tar-sync-vib、スペクトラル・シェイプS tar-sync(f)
及び残差成分R tar-sync(f)となる（ステップS1
6）。

【0074】また、図9に示すように、タイミングt =
t3~t4のターゲットフレームには、ターゲットフレ
ーム情報データINF tarが存在せず、シンクロナイゼー
ションモード = “2” となるが、タイミングt = t2~
t3のターゲットフレームには、ターゲットフレーム情
報データINF tarが存在するので、置換済ターゲット
フレーム情報データINF tar-syncをタイミングt = t
2~t3のターゲットフレームのターゲットフレーム情報
データである置換用ターゲットフレーム情報データI
NF holdとする処理P2を行い、以降の処理に用いられる
置換済ターゲットフレーム情報データINF tar-syncに
含まれるターゲット属性データは、平均アンプ静的成分
A tar-sync-sta、平均アンプビブラートの成分A tar-sy
nc-vib、ピッチ静的成分P tar-sync-sta、ピッチビブ
ラートの成分P tar-sync-vib、スペクトラル・シェイプS
tar-sync(f)及び残差成分R tar-sync(f)となる（ステッ

プS16）。

【0075】ステップS12の判別において、元無声/
有声検出信号U/V me(t)が無声（U）から有声（V）
に変化していない場合には（ステップS12；No）、
ターゲット無声/有声検出信号U/V tar(t)が有声
（V）から無声（U）に変化しているか否かを判別する
（ステップS13）。ステップS13の判別において、
ターゲット無声/有声検出信号U/V tar(t)が有声
（V）から無声（U）に変化している場合には（ステッ
プS13；Yes）、タイミングtの前のタイミング
t-1における元無声/有声検出信号U/V me(t-1)が有
声（V）かつターゲット無声/有声検出信号U/V tar
(t-1)が有声（V）であるか否かを判別する（ステップ
S19）。

【0076】例えば、図9に示すように、タイミングt
3においてターゲット無声/有声検出信号U/V tar(t)
が有声（V）から無声（U）に変化し、タイミングt-1
= t2~t3においては、元無声/有声検出信号U/V me
(t-1)が有声（V）かつターゲット無声/有声検出信号
U/V tar(t-1)が有声（U）となっている。ステップS
18の判別において、元無声/有声検出信号U/V me(t
-1)が有声（V）かつターゲット無声/有声検出信号U
/V tar(t-1)が有声（V）となっている場合には（ステ
ップS19；Yes）、当該ターゲットフレームには、
ターゲットフレーム情報データINF tarが存在しない
ので、シンクロナイゼーションモード = “2” とし、置
換用のターゲットフレーム情報データINF holdを当該
ターゲットフレームの前方向（forward）に存在するフ
レームのターゲットフレーム情報とする。

【0077】例えば、図9に示すように、タイミングt
= t3~t4のターゲットフレームには、ターゲットフレ
ーム情報データINF tarが存在しないので、シンクロ
ナイゼーションモード = “2” とし、置換用ターゲット
フレーム情報データINF holdを当該ターゲットフレ
ームの前方向に存在するフレーム（すなわち、タイミン
グt = t2~t3に存在するフレーム）のターゲットフレ
ーム情報データforwardとする。

【0078】そして、処理をステップS15に移行し、
シンクロナイゼーションモード = “0” であるか否かを
判別して（ステップS15）、以下、同様の処理を行
う。ステップS13の判別において、ターゲット無声/
有声検出信号U/V tar(t)が有声（V）から無声（U）
に変化していない場合には（ステップS13；No）、
タイミングtにおける元無声/有声検出信号U/V me
(t)が有声（V）から無声（U）に変化し、あるいは、
ターゲット無声/有声検出信号U/V tar(t)が無声
（U）から有声（V）に変化しているか否かを判別する
（ステップS14）。

【0079】ステップS14の判別において、タイミン
グtにおける元無声/有声検出信号U/V me(t)が有声

(V) から無声 (U) に変化し、かつ、ターゲット無声／有声検出信号 $U/V_{tar}(t)$ が無声 (U) から有声 (V) に変化している場合には (ステップ S14; Yes)、シンクロナイズーションモード = "0" とし、置換用ターゲットフレーム情報データ INF_{hold} を初期化 (clear) し、処理をステップ S15 に移行して、以下、同様の処理を行う。ステップ S14 の判別において、タイミング t における元無声／有声検出信号 $U/V_{me}(t)$ が有声 (V) から無声 (U) に変化せず、あるいは、ターゲット無声／有声検出信号 $U/V_{tar}(t)$ が無声 (U) から有声 (V) に変化していない場合には (ステップ S14; No)、そのまま処理をステップ S15 に移行し、以下同様の処理を行う。

【0080】 [3. 9] 正弦波成分属性データ選択部の動作

続いて、正弦波成分属性データ選択部 23 は、イージーシンクロナイズーション処理部 22 から入力された置換済ターゲットフレーム情報データ $INF_{tar-sync}$ に含まれるターゲット属性データのうち正弦波成分に関するターゲット属性データ (平均アンブ静的成分 $A_{tar-sync-sta}$ 、平均アンブビブラートの成分 $A_{tar-sync-vib}$ 、ピッチ静的成分 $P_{tar-sync-sta}$ 、ピッチビブラートの成分 $P_{tar-sync-vib}$ 及びスペクトラル・シェイプ $S_{tar-sync}(f)$) 及びコントローラ 29 から入力される正弦波成分属性データ選択情報に基づいて、新しい正弦波成分属性データである新規アンブ成分 A_{new} 、新規ピッチ成分 P_{new} 及び新規スペクトラル・シェイプ $S_{new}(f)$ を生成する。

【0081】 すなわち、新規アンブ成分 A_{new} については、次式により生成する。

$$A_{new} = A_{*-sta} + A_{*-vib} \quad (\text{ただし、*は、me又はtar-sync})$$

より具体的には、図 8 (D) に示すように、新規アンブ成分 A_{new} を元属性データの平均アンブ静的成分 A_{me-sta} あるいはターゲット属性データの平均アンブ静的成分 $A_{tar-sync-sta}$ のいずれか一方及び元属性データの平均アンブビブラートの成分 A_{me-vib} あるいはターゲット属性データの平均アンブビブラートの成分 $A_{tar-sync-vib}$ のいずれか一方の組み合わせとして生成する。

【0082】 また、新規ピッチ成分 P_{new} については、次式により生成する。

$$P_{new} = P_{*-sta} + P_{*-vib} \quad (\text{ただし、*は、me又はtar-sync})$$

より具体的には、図 8 (D) に示すように、新規ピッチ成分 P_{new} を元属性データのピッチ静的成分 P_{me-sta} あるいはターゲット属性データのピッチ静的成分 $P_{tar-sync-sta}$ のいずれか一方及び元属性データのピッチビブラートの成分 P_{me-vib} あるいはターゲット属性データのピッチビブラートの成分 $P_{tar-sync-vib}$ のいずれか一方の組み合わせとして生成する。

【0083】 また、新規スペクトラル・シェイプ $S_{new}(f)$ については、次式により生成する。

$S_{new}(f) = S * (f)$ (ただし、*は、me又はtar-sync) ところで、一般的にアンブ成分が大きい場合には、高域まで伸びた抜けの明るい音となり、アンブ成分が小さい場合には、逆にこもった音になる。そこで、新規スペクトラル・シェイプ $S_{new}(f)$ に関しては、このような状態をシミュレートすべく、図 11 に示すように、スペクトラル・シェイプの高域成分、すなわち、高域成分部分のスペクトラル・シェイプの傾きを新規アンブ成分 A_{new} の大きさに応じて補償するスペクトラルチルト補償 (spectral tilt correction) を行って、コントロールすることにより、よりリアルな音声を再生することができる。

【0084】 続いて、生成された新規アンブ成分 A_{new} 、新規ピッチ成分 P_{new} 及び新規スペクトラル・シェイプ $S_{new}(f)$ について、必要に応じてコントローラ 29 から入力される正弦波成分属性データ変形情報に基づいて、属性データ変形部 24 によりさらなる変形を行う。例えば、スペクトラル・シェイプを全体的に間延びさせる等の変形を行う。

【0085】 [3. 10] 残差成分選択部の動作 一方、残差成分選択部 25 は、イージーシンクロナイズーション処理部 22 から入力された置換済ターゲットフレーム情報データ $INF_{tar-sync}$ に含まれるターゲット属性データのうち残差成分に関するターゲット属性データ (残差成分 $R_{tar-sync}(f)$)、残差成分保持部 12 に保持されている残差成分信号 (周波数波形) $R_{me}(f)$ 及びコントローラ 29 から入力される残差成分属性データ選択情報に基づいて新しい残差成分属性データである新規残差成分 $R_{new}(f)$ を生成する。

【0086】 すなわち、新規残差成分 $R_{new}(f)$ については、次式により生成する。

$$R_{new}(f) = R * (f) \quad (\text{ただし、*は、me又はtar-sync})$$

この場合においては、me又はtar-syncのいずれを選択するかは、新規スペクトラル・シェイプ $S_{new}(f)$ と同一のものを選択するのがより好ましい。

【0087】 さらに、新規残差成分 $R_{new}(f)$ に関しても、新規スペクトラル・シェイプと同様な状態をシミュレートすべく、図 11 に示したように、残差成分の高域成分、すなわち、高域成分部分の残差成分の傾きを新規アンブ成分 A_{new} の大きさに応じて補償するスペクトラルチルト補償 (spectral tilt correction) を行って、コントロールすることにより、よりリアルな音声を再生することができる。

【0088】 [3. 11] 正弦波成分生成部の動作 続いて、正弦波成分生成部 26 は、属性データ変形部 24 から出力された変形を伴わない、あるいは、変形を伴う新規アンブ成分 A_{new} 、新規ピッチ成分 P_{new} 及び新規スペクトラル・シェイプ $S_{new}(f)$ に基づいて、当該フレ

ームにおける新たな正弦波成分 (F''_0 、 A''_0)、
(F''_1 、 A''_1)、(F''_2 、 A''_2)、……、
($F''_{(N-1)}$ 、 $A''_{(N-1)}$)のN個の正弦波成分(以下、
これらをまとめて F''_n 、 A''_n と表記する。 $n=0\sim$
($N-1$)。))を求める。

【0089】より具体的には、次式により新規周波数
 F''_n および新規アンプ A''_n を求める。

$$F''_n = F'_n \times P_{new}$$

$$A''_n = S_{new}(F''_n) \times A_{new}$$

なお、完全倍音構造のモデルとして捉えるのであれば、

$$F''_n = (n+1) \times P_{new}$$

となる。

【0090】[3.12] 正弦波成分変形部の動作
さらに、求めた新規周波数 F''_n および新規アンプ A''_n
について、必要に応じてコントローラ29から入力さ
れる正弦波成分変形情報に基づいて、正弦波成分変形部
27によりさらなる変形を行う。例えば、偶数倍音成分
の新規アンプ A''_n (= A''_0 、 A''_2 、 A''_4 、…
…)だけを大きく(例えば、2倍する)等の変形を行
う。これによって得られる変換音声にさらにバラエティ
を持たせることが可能となる。

【0091】[3.13] 逆高速フーリエ変換部の動作

次に逆高速フーリエ変換部28は、求めた新規周波数
 F''_n および新規アンプ A''_n (=新規正弦波成分)並
びに新規残差成分 $R_{new}(f)$ をFFTバッファに格納し、
順次逆FFTを行い、さらに得られた時間軸信号を一部
重複するようにオーバーラップ処理し、それらを加算す
る加算処理を行うことにより新しい有声音の時間軸信号
である変換音声信号を生成する。

【0092】このとき、コントローラ29から入力され
る正弦波成分/残差成分バランス制御信号に基づいて、
正弦波成分及び残差成分の混合比率を制御し、よりリア
ルな有声音号を得る。この場合において、一般的には、
残差成分の混合比率を大きくするとざらついた声を得ら
れる。この場合において、FFTバッファに新規周波数
 F''_n および新規アンプ A''_n (=新規正弦波成分)並
びに新規残差成分 $R_{new}(f)$ を格納するに際し、異なるビ
ッチ、かつ、適当なビッチで変換された正弦波成分をさ
らに加えることにより変換音声信号としてハーモニーを
得ることができる。さらにシーケンサ31により伴奏音
に適合したハーモニービッチを与えることにより、伴奏
に適合した音楽的ハーモニーを得ることができる。

【0093】[3.14] クロスフェーダの動作

次にクロスフェーダ部30は、元無声/有聲検出信号U
/Vme(t)に基づいて、入力音声信号Svが無声(U)で
ある場合には、入力音声信号Svをそのままミキサ33
に出力する。また、入力音声信号Svが有聲(V)であ
る場合には、逆高速フーリエ変換部28が出力した変換
音声信号をミキサ33に出力する。この場合において、

切替スイッチとしてクロスフェーダ部30を用いている
のは、クロスフェード動作を行わせることによりスイ
ッチ切替時のクリック音の発生を防止するためである。

【0094】[3.15] シーケンサ、音源部、ミキ
サ及び出力部の動作

一方、シーケンサ31は、カラオケの伴奏音を発生す
るための音源制御情報を例えば、MIDI (Musical Inst
rument Digital Interface) データなどとして音源部3
2に出力する。これにより音源部32は、音源制御情報
に基づいて伴奏信号を生成し、ミキサ33に出力する。
ミキサ33は、入力音声信号Svあるいは変換音声信号
のいずれか一方及び伴奏信号を混合し、混合信号を出力
部34に出力する。出力部34は、図示しない増幅器を
有し混合信号を増幅して音響信号として出力することと
なる。

【0095】[4] 実施形態の変形例

[4.1] 第1変形例

以上の説明においては、属性データとしては、元属性デ
ータあるいはターゲット属性データのいずれかを選択的
に用いる構成としていたが、元属性データ及びターゲッ
ト属性データの双方を用い、補間処理を行うことにより
中間的な属性を有する変換音声信号を得るように構成す
ることも可能である。しかしながら、このような構成に
よれば、ものまねをしようとする歌唱者及びものまねの
対象(target)となる歌唱者のいずれにも似ていない変
換音声信号が得られる場合もある。また、特にスペクトル
・シェイプを補間処理によって求めた場合には、ものま
ねをしようとする歌唱者が「あ」を発音し、ものまねの
対象となる歌唱者が「い」を発音している場合などに
は、「あ」でも「い」でもない音が変換音声として出力
される可能性があり、その取扱には注意が必要である。

【0096】[4.2] 第2変形例

正弦波成分の抽出は、この実施形態で用いた方法に限ら
ない。要は、音声信号に含まれる正弦波を抽出できれば
よい。

【0097】[4.3] 第3変形例

本実施形態においては、ターゲットの正弦波成分及び残
差成分を記憶したが、これに換えて、ターゲットの音声
そのものを記憶し、それを読み出してリアルタイム処理
によって正弦波成分と残差成分とを抽出してもよい。す
なわち、本実施形態でもものまねをしようとする歌唱者の
音声に対して行った処理と同様の処理をターゲットの歌
唱者の音声に対して行ってもよい。

【0098】[4.4] 第4変形例

本実施形態においては、属性データとして、ビッチ、ア
ンプ、スペクトル・シェイプの全てを取り扱ったが、
少なくともいずれか一つを扱うようにすることも可能で
ある。

【0099】[4.5] 第5変形例

本実施形態の補間合成部9におけるデータ構成は、その

他の各部（例えばピッチ正規化部17～正弦波成分変形部27に至る区間）におけるデータ構成としても良いことは言うまでもない。特に、ターゲットフレーム情報保持部20においては、1曲分のターゲットフレーム情報が記憶されるため、上記データ構成を用いることによるデータ量の削減効果が大きい。

【0100】[5] 実施形態の効果

以上のように、本実施形態によれば、周波数 F_k に対して差分値 dF_k 、比率 rF_k または対数値 cF_k を記憶するようにしたため、僅かな記憶容量で高精度な音声特徴情報を記憶できる。さらに、アンプ A_k に対して1バイト中の7ビットを割り当てて128dBのダイナミックレンジを確保するとともに残りの1ビットにおいて周波数を特定するための情報（上記対数値 cF_k 等）が存在するか否かを示す態様においては、さらに記憶容量を削減することができる。また、アンプ A_k に代えて「 $nA_k = \alpha \cdot (A_k + \beta)$ 」によるアンプ nA_k を記憶することにより、所望のダイナミックレンジに応じて可能な限り高い分解能を確保することが可能である。

【0101】

【発明の効果】以上説明したように本発明によれば、各乗算結果 $F0 \times k$ と各周波数 F_k との差分または割合を記憶することによって音声の特徴を記憶するから、僅かな記憶容量で高精度な音声特徴情報を記憶できる。

【図面の簡単な説明】

【図1】 本発明の一実施形態の構成を示すブロック図（その1）である。

【図2】 本発明の一実施形態の構成を示すブロック図（その2）である。

【図3】 実施形態におけるフレームの状態を示す図である。

【図4】 実施形態における周波数スペクトルのピーク*

* 検出を説明するための説明図である。

【図5】 実施形態におけるフレーム毎のピーク値の連携を示す図である。

【図6】 実施形態における周波数値の変化状態を示す図である。

【図7】 実施形態における処理過程における確定成分の変化状態を示す図である。

【図8】 実施形態における信号処理の説明図である。

【図9】 イーजीシンクロナイゼーション処理のタイミングチャートである。

【図10】 イーजीシンクロナイゼーション処理フローチャートである。

【図11】 スペクトラル・シェイプのスペクトラルルト補償について説明する図である。

【符号の説明】

1…マイク、2…分析窓生成部、3…入力音声信号切出部、4…高速フーリエ変換部、5…ピーク検出部、6…無声/有声検出部、7…ピッチ抽出部、8…ピーク連携部、9…補間合成部、10…残差成分検出部、11…高速フーリエ変換部、12…残差成分保持部、13…正弦波成分保持部、14…平均アンプ演算部、15…アンプ正規化部、16…スペクトラル・シェイプ演算部、17…ピッチ正規化部、18…元フレーム情報保持部、19…静的変化/ビブラートの変化分離部、20…ターゲットフレーム情報保持部、21…キーコントロール/テンポチェンジ部、22…イーजीシンクロナイゼーション処理部、23…正弦波成分属性データ選択部、24…属性データ変形部、25…残差成分選択部、26…正弦波成分生成部、27…正弦波成分変形部、28…逆高速フーリエ変換部、29…コントローラ、30…クロスフェーダ部、31…シーケンサ、32…音源部、33…ミキサ、34…出力部。

Fig.3

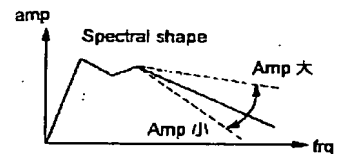
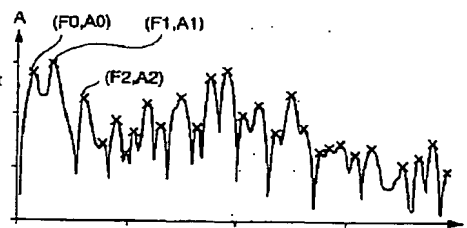
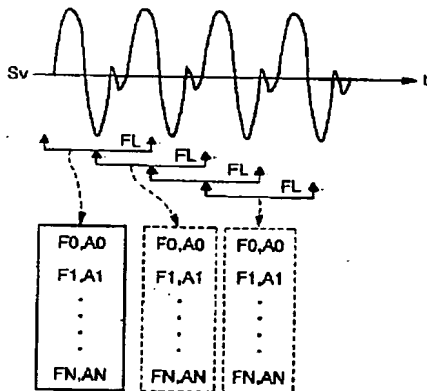
【図3】

Fig.4

【図4】

Fig.11

【図11】



【図1】

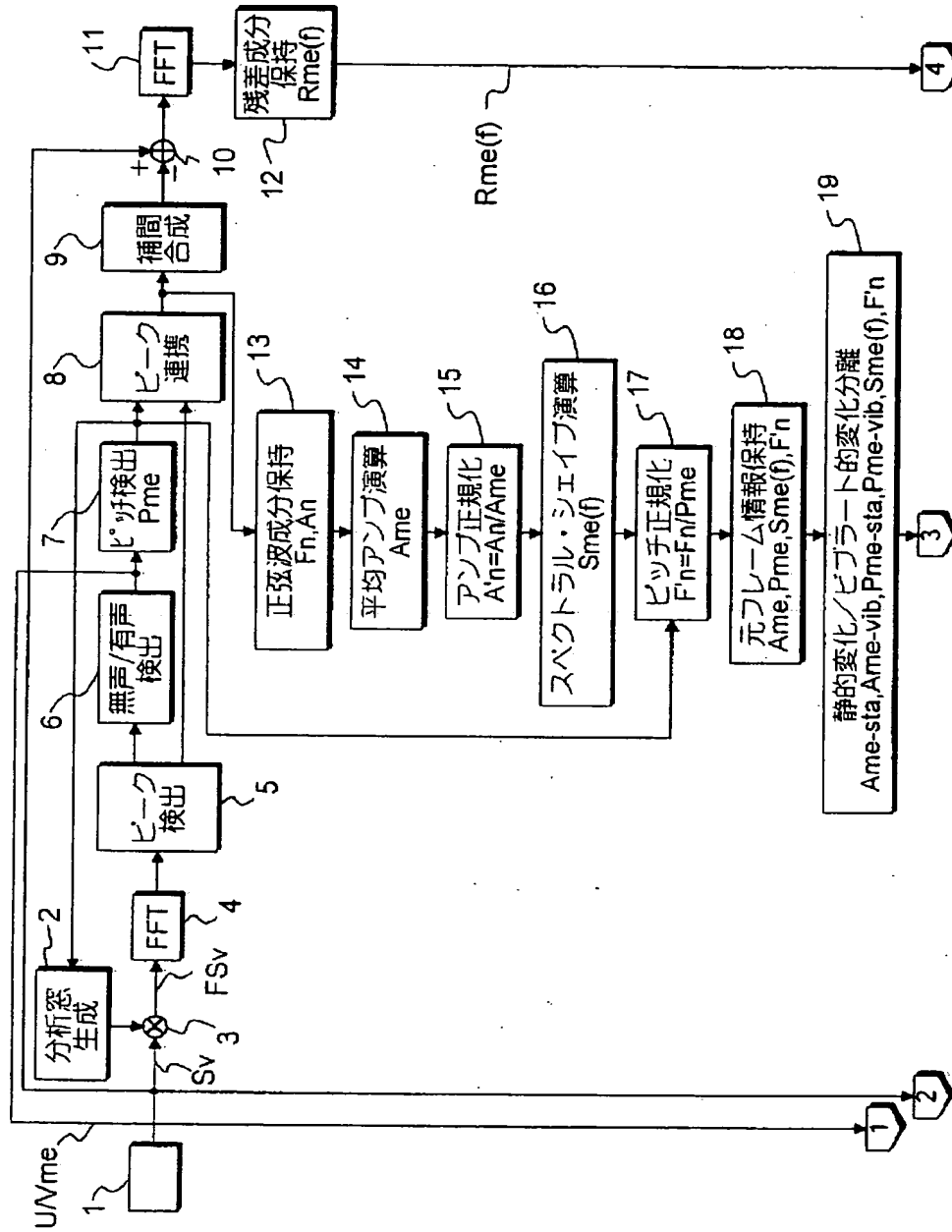


Fig. 2

【図2】

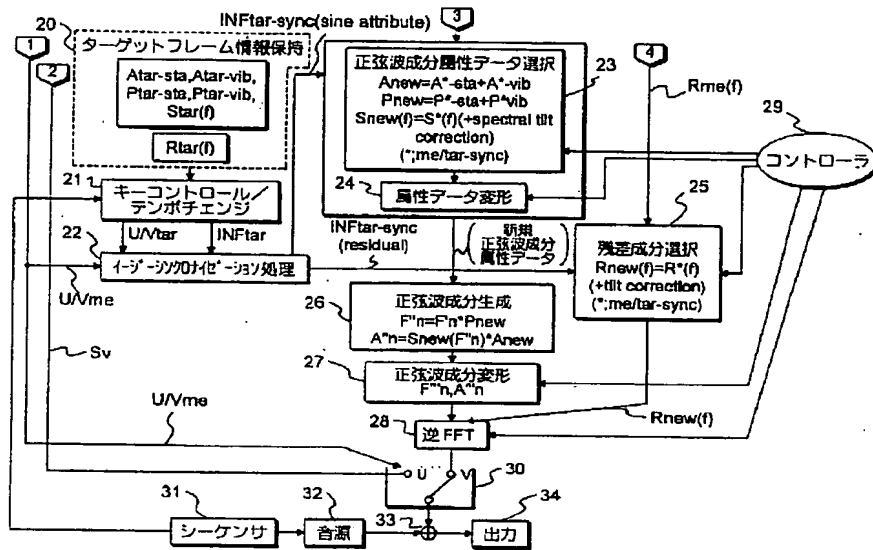


Fig. 5

【図5】

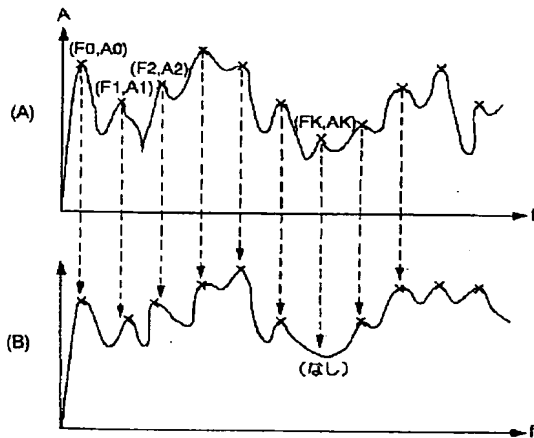


Fig. 6

【図6】

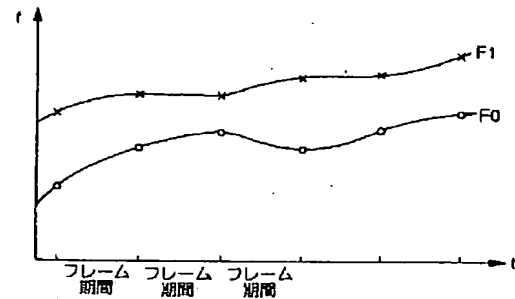


Fig. 7

【図7】

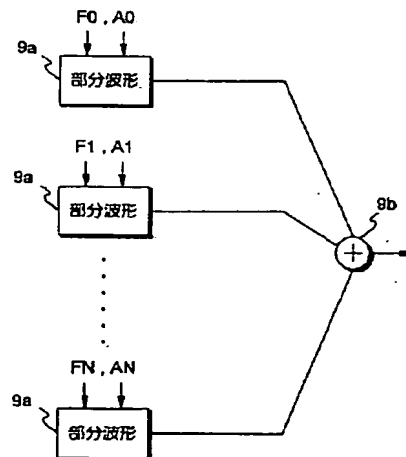


Fig. 8

【図8】

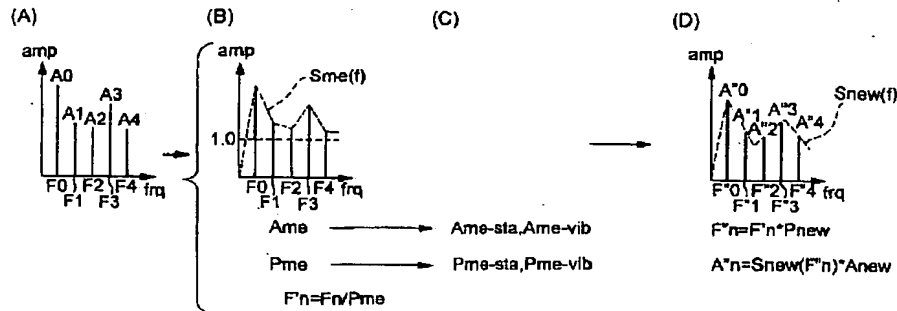


Fig. 9

【図9】

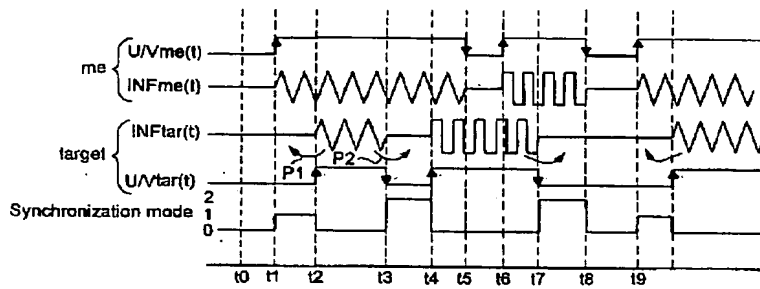
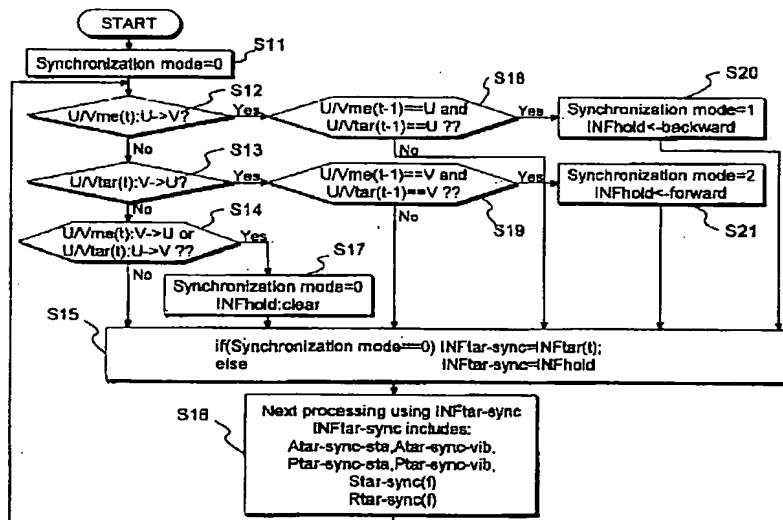


Fig. 10

【図10】



**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☐ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☒ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☒ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.